university of groningen

University Medical Center Groningen

University of Groningen

Reachability of Five Gossip Protocols

Van Ditmarsch, Hans; Gattinger, Malvin; Kokkinis, Ioannis; Kuijer, Louwe B.

Published in:
Reachability Problems

DOI:
10.1007/978-3-030-30806-3_17

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Final author's version (accepted by publisher, after peer review)

Publication date:
2019

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Van Ditmarsch, H., Gattinger, M., Kokkinis, I., & Kuijer, L. B. (2019). Reachability of Five Gossip Protocols. In E. Filiot, R. Jungers, & I. Potapov (Eds.), *Reachability Problems: RP 2019: International Conference on Reachability Problems* (pp. 218-231). (Lecture Notes in Computer Science; Vol. 11674). Springer. https://doi.org/10.1007/978-3-030-30806-3_17

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 06-05-2023

Reachability of Five Gossip Protocols

Hans van Ditmarsch[1][0000−0003−4526−8687],
Malvin Gattinger[2][0000−0002−2498−5073],
Ioannis Kokkinis[3][0000−0001−7521−0553], and Louwe B. Kuijer[4]

[1] CNRS, LORIA, University of Lorraine, France · hans.van-ditmarsch@loria.fr
[2] University of Groningen, The Netherlands · malvin@w4eg.eu
[3] TU Dortmund, Germany · ioannis.kokkinis@tu-dortmund.de
[4] University of Liverpool, United Kingdom · l.b.kuijer@gmail.com

Abstract Gossip protocols use point-to-point communication to spread information within a network until every agent knows everything. Each agent starts with her own piece of information ('secret') and in each call two agents will exchange all secrets they currently know. Depending on the protocol, this leads to different distributions of secrets among the agents during its execution. We investigate which distributions of secrets are *reachable* when using several distributed epistemic gossip protocols from the literature. Surprisingly, a protocol may reach the distribution where all agents know all secrets, but not all other distributions. The five protocols we consider are called ANY, LNS, CO, TOK, and SPI. We find that TOK and ANY reach the same distributions but all other protocols reach different sets of distributions, with some inclusions. Additionally, we show that all distributions are *subreachable* with all five protocols: any distribution can be reached, if there are enough additional agents.

Keywords: Gossip · Networks · Reachability

1 Introduction

Let each of a set of agents $\{a, b, c, \ldots\}$ know a single secret $\{A, B, C, \ldots\}$, respectively. The agents can communicate via telephone calls. When they call, they share all the secrets they know at the moment the call takes place. An agent who knows all secrets is an *expert*. The goal is to turn all agents into experts. A protocol to achieve this state of knowledge is called a *gossip protocol* [10,11].

Here we consider five gossip protocols of a distributed nature [1,2,3,4]:

ANY *Any call is allowed, i.e., for every two agents a and b, a is allowed to call b.*

CO *Short for "call once". An agent a may call b iff they have not spoken before, i.e., if a has not called b before and b has not called a before.*

LNS *Short for "learn new secrets". Agent a may call b iff a doesn't yet know B.* During the execution of this protocol after every call at least one new secret is learned, hence the protocol name.

TOK *Short for "token". Agent a may call b iff a has a token. The caller passes her token to the callee.*

Every agent starts with a token. After a call all tokens held by an agent merge to one. In this protocol an agent who lost her token can get it back when she receives a new call.

Equivalently, we can say that a may call b iff a has either not been involved in any calls, or a was the callee in the last call a was involved in.

SPI *Short for "spider". Agent a may call b iff a has a token. The callee passes her token to the caller.*

Every agent starts with a token. After a call all tokens held by an agent merge to one. In this protocol an agent who has been called loses her token permanently and can never initiate a call again. This protocol tends to lead to a small number of agents making many calls. When drawn as a graph, this looks like a spider web with the agent making the calls at the centre, hence the name "spider".

Equivalently, we can say that a may call b iff a has never received any calls.

All protocols run in a sequential manner as follows: starting from the situation where each agent only knows her own secret, each moment in time a single call satisfying the protocol condition is selected and executed. The selection of calls continues until all agents are experts. Here we investigate which distributions of secrets may be reached during the protocol execution (under *any sequence* of calls), hence we do not have to fix a specific algorithm for call selection.

Knowing which distribution can be reached by which protocol can help the agents (or an external observer) understand which protocol is being used during the exchange of information. Moreover, reachability can be of importance for security or privacy reasons.

Let us illustrate the topic of reachability by an example with three agents. We represent a distribution of secrets by listing the secrets known by each agent. Given initial distribution (A, B, C), the call ab (the call from a to b) results in (AB, AB, C). (Strictly, we go from $(\{A\}, \{B\}, \{C\})$ to $(\{A, B\}, \{A, B\}, \{C\})$.) This is therefore ANY-reachable. After the call sequence $ab; bc; ac$, which is permitted in ANY, LNS, and CO, all three agents are experts. But already for three agents there is a difference between the five protocols. The sequence $ab; bc; ca$ is CO-permitted but not LNS-permitted: as c already knows A, the call ca is not allowed in LNS. The sequence $ab; bc; ab$ is not CO-permitted (repeating ab is not allowed); and clearly if a call is not CO-permitted it is also not LNS-permitted. Call sequence $ab; ba; ab$ is TOK-permitted but not SPI-permitted, whereas call sequence $ab; ab$ is SPI-permitted but not TOK-permitted.

We assume that communication between all agents is possible, i.e., a complete network topology. Unreachability results in this setting are very strong: if one of the five protocols cannot reach a distribution s assuming a complete network topology, then this protocol also cannot reach s assuming any other topology. It is not difficult to see that already for two agents, unreachable distributions can occur: the distribution (AB, A) cannot be reached by any of the five protocols.

We also study a less restrictive notion called *subreachability*. Given agents a, b, c, if b calls c, and then c calls a, the resulting distribution is (ABC, BC, ABC).

The restriction of that distribution to the agents a and b only is (AB, B). We say that distribution (AB, B), although not reachable, is subreachable. Knowing the knowledge situations that can be subreached by a protocol is particularly interesting when the number of agents is not common knowledge among the agents, or when the agents have limited reasoning power and cannot reason like "there are two agents beside me and a call has taken place, so these agents now know each other's secrets". In such a situation the agents should not only consider the reachable but also the subreachable distributions possible.

We further investigate reachability under *unordered distributions* ("given n agents and n subsets of the set of all secrets, is there a bijection between these sets of agents and subsets?"). Unordered distributions should be taken into consideration by an observer who is uncertain about which agent holds which sets of secrets in a distribution.

Our Contributions. For up to three agents all five protocols can reach the same distributions. Thus, with at most three agents present, an observer cannot tell which protocol is currently used, by simply observing the distributions of secrets. But with four or more agents there is a difference in the reachability strength. In Figure 1 we give a complete overview of each protocol's reachability strength. Figure 1 (together with the relevant theorems) can serve as guide for an (internal or external observer) that wants to know which protocol the agents are using for information exchange. For example, if the observer finds out that the distribution of secrets $(ABCD, ABCD, ABC, ABD)$ has appeared, then she can be certain that agents are not using the CO-protocol since this distribution is not CO-reachable (see Theorem 3).

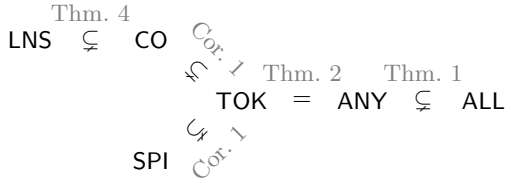

Figure 1. Overview of results, in which a protocol's name stands for the set of distributions reachable by it and ALL stands for the set of all distributions. Besides transitivity no other inclusions hold, i.e., SPI and CO properly intersect (see Theorem 3 and Corollary 1), and so do SPI and LNS (see Theorems 4 and 5).

In Theorem 6 we show that all distributions are subreachable by all five protocols. As we mentioned before this has the consequence that an observer cannot infer which of the five protocols is being used if, for example, she does not know how many agents there are, since every distribution of secrets can possibly occur among a subset of the agents using any of the five protocols. Finally in Theorem 7 we show that SPI, ANY and TOK reach the same set of unordered

distributions. The consequence of the latter theorem is that the observer cannot distinguish between the protocols SPI, ANY and TOK if she is uncertain about which agent holds which set of secrets in a given distribution.

Related Work. The combinatorial properties of gossip protocols have been investigated several times in the literature. In [1] the focus is on distributed gossip, including information change in one direction only, and termination. The extension (permitted call sequences of the protocols) and the characterization of the classes of graphs where the (dynamic versions) of our protocols terminate were investigated in [5], where their main result is for LNS (in [9] the same question was answered for the (static) protocol CO). In [2,3] the focus is on the logical dynamics of call exchange. In [6,7] the gossip protocols were treated as random processes and it was shown that TOK and SPI have the same expectation. As simulations (some of which where theoretically corroborated) in [6,7] indicate, the expected duration of all protocols considered here is of the order $n \log n$, the 'usual' suspect in the gossiping community, but the constant factor may be different.

Organization of the Paper. In Section 2 we present all the definitions and relevant notions that are necessary for understanding our results. In Section 3 we present our main result which is the comparison of the reachability strength of the 5 protocols. In Section 4 we study the subreachability strength of our protocols and their reachability strength in unordered distributions. Finally in Section 5 we give directions for further work, mainly on parallel calls.

2 Terminology for Gossip Protocols and Reachability

In this section we give formal definitions for the notions of secret (sub)distributions and (sub)reachability. We always assume a complete network topology. A set of agents is represented by $\mathcal{A}$. We use the lower-case letters $a, b, c, d, \ldots$ for agents. At the start of any gossip protocol each agent has a unique secret. We denote the secrets by the corresponding upper-case letters $A, B, C, D, \ldots$ and there are no other secrets.

Definition 1 (Distribution of Secrets). *An* n-distribution of secrets *for a set of agents* $\mathcal{A} = \{a_1, \ldots, a_n\}$ *is an ordered* n*-tuple* $(S_{a_1}, \ldots, S_{a_n})$ *where each* S_{a_i} *is a subset of the set of all secrets* $\{A_1, \ldots, A_n\}$*. In the* initial distribution *every agent knows only her own secret, i.e.* $S_{a_i} = \{A_i\}$ *for all* a_i*. An agent* a_i *is an* expert *iff she knows all secrets, i.e. iff* $S_{a_i} = \{A_1, \ldots, A_n\}$*. In the* final distribution *every agent is an expert.*

In general, a distribution $(S_{a_1}, \ldots, S_{a_n})$ represents the situation in which each agent a_i knows exactly the secrets in S_{a_i}. We drop the references to $\mathcal{A}$, n and secrets if this causes no confusion. We write (ABC, AB, ABC) instead of $(\{A, B, C\}, \{A, B\}, \{A, B, C\})$. We use the letters s, t (possibly primed or with subscripts) to represent a distribution. Finally, we observe that a distribution of secrets implicitly assumes an ordering on the agents.

Definition 2 (Call). *A* call *is an ordered pair (a, b), where $a \neq b$ for some agents a, b. We write ab instead of (a, b). A* call sequence *is a (possibly empty) finite or infinite sequence of calls. We write $ab; cd; \ldots$ for a call sequence. If ab occurs in a call sequence σ, we also write $ab \in \sigma$, slightly abusing language. By (ab) we mean the call ab or the call ba. Let $s = (S_{a_1}, \ldots, S_{a_n})$ be a distribution and consider $(a_i a_j)$ for some $i < j$. We apply any of the two calls $a_i a_j$ and $a_j a_i$ to $(S_{a_1}, \ldots, S_{a_n})$ as follows and obtain the new distribution*

$$s^{a_i a_j} := s^{a_j a_i} :=$$
$$(S_{a_1}, \ldots, S_{a_{i-1}}, S_{a_i} \cup S_{a_j}, S_{a_{i+1}}, \ldots, S_{a_{j-1}}, S_{a_i} \cup S_{a_j}, S_{a_{j+1}}, \ldots, S_{a_n}) \ .$$

We apply a finite call sequence σ to a distribution s as follows:

$$s^{\sigma} := \begin{cases} s, & \text{if } \sigma = \epsilon \\ (s^{ab})^{\tau}, & \text{if } \sigma = ab; \tau \ . \end{cases}$$

For example, we have $(A, B, C)^{ab;bc} = (AB, AB, C)^{bc} = (AB, ABC, ABC)$.

A call sequence σ is P-*permitted* if the restrictions of P allow every call in σ to be executed in the order given in σ. A P-permitted call sequence will also be called P-call sequence. A call sequence σ is called *successful* if the application of σ to an initial distribution leads to the final distribution where all agents know all secrets. If the applications of either σ or τ to the initial distribution lead to the same distribution we write $\sigma \approx \tau$.

Definition 3 (Reachability). *A distribution s is* P-reachable *if s can be obtained by applying a P-permitted call sequence on the initial distribution.*

The ANY-permitted calls are also called the *possible* calls and an ANY-reachable distribution is also called a *possible* distribution.

From a given n-distribution we can derive the set of possible calls that could have contributed to reaching that distribution, including an order on their execution. It is defined as follows.

Definition 4. *Let s be a distribution. The set of potential calls for s is $PC(s) := \{ab \mid A \in S_b \text{ and } B \in S_a, \text{ for some agents } a, b\}$. The order $<_s$ on $PC(s)$ is defined as follows. For any $ab, cd \in PC(s)$:*

$$\begin{aligned} ab < cd \ \text{ if } \ & a = c \text{ and } D \notin S_b, \text{ or} \\ & b = c \text{ and } D \notin S_a, \text{ or} \\ & a = d \text{ and } C \notin S_b, \text{ or} \\ & b = d \text{ and } C \notin S_a. \end{aligned}$$

The pair $(PC(s), <_s)$ is called the set of potential call sequences *(for s).*

A call sequence σ consisting of calls from $PC(s)$ respects *the order $<_s$ if, for every $ab <_s cd$, no occurrence of (ab) in σ is after any occurrence of (cd) in σ.*

Let $(ab) <_s (cd)$ denote: $ab <_s cd$, $ba <_s cd$, $ab <_s dc$, and $ba <_s dc$. Now let σ and τ be call sequences. By $\sigma <_s \tau$ we mean that for every $xy \in \sigma$ and

every $zw \in \tau$ if xy is related to zw then $xy <_s zw$; and that no pair of calls in σ are comparable and that the same holds for τ. We may additionally employ $(\sigma) <_s (\tau)$ meaning that for every $xy \in \sigma$ and every $zw \in \tau$, if xy and zw are comparable then $(xy) <_s (zw)$.

The proof of the next proposition is obvious.

Proposition 1. *Each distribution s uniquely determines a pair $(PC(s), <_s)$. Distribution s can only be obtained by a call sequence in which only calls in $PC(s)$ occur, and that respects the order $<_s$.*

We note (i) that a pair $(PC(s), <_s)$ does *not* uniquely determine a given distribution s, (ii) that calls may occur more than once (for example, in both directions, and as long as the order $<_s$ is respected), and (iii) that not all calls in $PC(s)$ need occur in a sequence reaching s. The proof of Theorem 5 demonstrates (i) and (iii). Concerning (ii), note that any call ab can be followed by (if the protocol so permits) the dual call ba as long as neither a nor b have been involved in other calls, without the second call ba affecting the distribution at that time.

Example 1. Consider the 4-distribution $s = (ABCD, ABCD, ABCD, ABCD)$. We have $PC(s) = \{ab, ac, ad, bc, bd, cd, ba, ca, da, cb, db, dc\}$ and $<_s = \emptyset$. Two different call sequences reaching s are $ab; cd; ac; bd$ and $ac; bd; ab; cd$. There are also call sequences that respect $<_s$ and do not reach s (e.g. $ab; ac; bd; cd$).

As a second example, consider the 3-distribution $t = (AB, ABC, BC)$. Then $PC(t) = \{ab, bc, ba, cb\}$ and $(ab) <_t (bc)$ and $(bc) <_t (ab)$. No call sequence respecting $<_t$ reaches t. Indeed, t is not ANY-reachable.

Finally we present the notion of *subreachability* that uses that of the *restriction* of a distribution.

Definition 5. *Suppose we have $\mathcal{A}' \subseteq \mathcal{A}$ with $m = |\mathcal{A}'|$ and $n = |\mathcal{A}|$. The* $\mathcal{A}'$-restriction *of an n-distribution $(S_{a_1}, \ldots, S_{a_n})$ for $\mathcal{A}$ is the m-distribution $(S_{a'_1}, \ldots, S_{a'_m})$ such that for all $a'_j \in \mathcal{A}'$, if $a_i = a'_j$ then $S_{a'_j} = \{A_k \in S_{a_i} \mid a_k \in \mathcal{A}'\}$.*

Definition 6 (Subreachability). *A distribution s for a set of agents $\mathcal{A}'$ is* P-subreachable *if there is a distribution t for an extended set of agents $\mathcal{A} \supseteq \mathcal{A}'$ such that t is* P*-reachable and s is the $\mathcal{A}'$-restriction of t.*

Note that P-reachable implies P-subreachable, namely when $\mathcal{A}' = \mathcal{A}$ above.

3 Reachability

In this section we provide an answer to the question: "are all P_1-reachable distributions also P_2-reachable?" for any P_1 and P_2 from the five protocols. It is interesting that although all five protocols can reach the final distribution on complete graphs [5], their reachability strength on intermediate distributions varies.

Theorem 1.

1. *There is a distribution that is not reachable by any of the five protocols.*
2. *Every* CO-*,* LNS-*,* SPI- *and* TOK-*distribution is* ANY-*reachable.*
3. *Every* LNS-*reachable distribution is* CO-*reachable.*

Proof. This follows from the protocol definitions and because (AB, A) is not reachable by any of the protocols.

Our next, rather unexpected, result is that, although TOK has a stricter calling condition than ANY, these two protocols reach the same set of distributions. Recall that TOK can be thought of as demanding that, in order to make a call, an agent has to possess a token. Every agent starts out with a token, and in a call ab the token of a is given to b. In the following lemma we use the fact that a call ab can be followed by a call ba in which the token is returned to a.

Lemma 1 (Token Density Lemma). *Let s be a* TOK-*reachable distribution and let a, b be two agents. Then s can be reached by a* TOK-*call sequence σ such that after the execution of σ at least one of a and b have a token.*

Proof. The Lemma follows easily from the following more general claim.

Claim. Let σ be any TOK sequence, let $k \in \mathbb{N}$, $I = \{1, \ldots, k\}$ and let $f, g : I \to \mathcal{A}$ be injections such that $f(I) \cap g(I) = \emptyset$ for some set of agents $\mathcal{A}$. Then there is a TOK sequence σ' such that (i) $\sigma \approx \sigma'$ and (ii) for every $1 \leq i \leq k$ at least one of $f(i)$ and $g(i)$ has a token after σ'.

Proof (of the Claim and the Lemma). By induction on the length of σ. If σ is of length 1 the claim is trivial. Assume then as induction hypothesis that the claim holds for all sequences shorter than σ. Now, let $\sigma = \tau; ab$. We distinguish whether the agents of the final call in σ are in the images of f and g.

- Suppose $a, b \notin f(I) \cup g(I)$. Then let $f', g' : I \cup \{k+1\} \to \mathcal{A}$ be extensions of f and g with $f'(k+1) = a, g'(k+1) = b$. By the induction hypothesis, there is τ' such that $\tau \approx \tau'$ and for every $1 \leq i \leq k+1$ either $f(i)$ or $g(i)$ has a token after τ'. Then $\tau'; (ab) \approx \sigma$ and for every $1 \leq i \leq k$, either $f(i)$ or $g(i)$ has a token after $\tau'; (ab)$.
- Suppose $a \in f(I) \cup g(I)$ and $b \notin f(I) \cup g(I)$. Without loss of generality, suppose that $f(1) = a$. Now, let f', g' be as f, g except $g'(1) = b$. By the induction hypothesis, there is a τ' such that $\tau \approx \tau'$ and either $f'(i)$ or $g'(i)$ ends up with a token. In particular, either a or b has a token after τ'. If a has the token, let $\sigma' = \tau'; ab; ba$, otherwise let $\sigma' = \tau'; ba$. In either case, (i) $\sigma' \approx \sigma$, (ii) for $i > 1$ either $f(i)$ or $g(i)$ has a token because they had it after τ' and (iii) a has a token so either $f(1)$ or $g(1)$ has a token.
- Suppose $a = f(i)$ and $b = g(i)$. By the induction hypothesis τ' exists, and $\sigma' = \tau'; (ab)$ suffices.

– Suppose $a = f(i)$ and $b \in f(I) \cup g(I) \setminus g(i)$. Without loss of generality, $b = f(j)$. Let f', g' be as f, g except $g'(i) = b$ and $f'(j) = g(i)$. Let τ' be such that $\tau \approx \tau'$ and for every l either $f'(l)$ or $g'(l)$ ends up with a token. Since $f'(i) = a$ and $g'(i) = b$, the sequence $\tau'; (ab)$ is TOK. Note furthermore that $f'(j) = g(i)$ and $g'(j) = g(j)$, so one of the pairs $(a, g(i))$ and $(f(j), b)$ has at least one token. By inverting the (ab) call if necessary, we can ensure that the other pair keeps the token of the (ab) call. As such, either $\tau'; (ab); (ba)$ or $\tau'; (ab)$ satisfies the conditions of the claim. □

Theorem 2. *Every* ANY*-reachable distribution is* TOK*-reachable.*

Proof. We will show that for every ANY sequence σ there is a TOK sequence σ' such that $\sigma \approx \sigma'$. The proof proceeds by induction on the length of the call sequence σ and by repeatedly applying Lemma 1.

If σ is of length 1, then σ is a TOK sequence. Assume then as induction hypothesis that the theorem holds for all sequences shorter than σ, and let $\sigma = \tau; ab$. By the induction hypothesis, there is a TOK sequence τ' such that $\tau \approx \tau'$. Because τ' is a TOK sequence it follows from Lemma 1 that there is a TOK sequence τ'' such that (i) $\tau' \approx \tau''$ and (ii) either a or b has a token after τ''. It follows that $\sigma' = \tau''; (ab)$ is a TOK sequence, and $\sigma \approx \sigma'$. □

We continue to compare the sets of distributions reachable by all other protocols. Theorems 3 and 4 are generalized versions of [6, Theorems 3 and 4].

Theorem 3. *There is a* SPI*-reachable distribution that is not* CO*-reachable.*

Proof. Consider the 4-distribution $t = (ABCD, ABCD, ABC, ABD)$. We show that in order to reach t one has to choose the same call twice.

– The initial 4-distribution is (A, B, C, D).
– Since c and d must not learn each others secret, the first call cannot be cd. Furthermore, if the first call is ac then, when d learns a's secret, she will also learn c's secret. With similar arguments we can show that the first call cannot be ad, bc or bd. Thus in order to reach t we have to select ab which leads to (AB, AB, C, D).
– Now, d has to learn A and B. So, without loss of generality the next call is ad which leads to (ABD, AB, C, ABD).
– Now, c has to learn A and B. The only way of achieving this is by selecting cb which leads to (ABD, ABC, ABC, ABD).
– Until now we have made the CO-permitted call sequence: $ab; ad; bc$. The only way of reaching t is by selecting call ab again, which is a violation of CO.

The call sequence that reached t is: $\sigma = ab; ad; cb; ab$. No agent who has been called initiates a call, hence σ is SPI-permitted. □

Theorem 4. *There is a* CO*- and* SPI*-reachable distribution that is not* LNS*-reachable.*

Proof. Consider the 6-distribution:

$$t = (ABCDEF, ABC, ABCDE, ABCDEF, DEF, ABDEF) \ .$$

We will show that we can reach t without violating CO or SPI, but at the price of having to make a call between agents that already know each other's secrets.

- The initial 6-distribution is (A, B, C, D, E, F).
- Agent b has to learn A and C and nothing else and e has to learn d and f and nothing else. So, without loss of generality, the first four calls can be $ab; cb; ed; ef$, which are clearly both SPI-and CO-permitted and lead to

$$(AB, ABC, ABC, DE, DEF, DEF) \ .$$

- Now c has to learn everything but F. The only way of achieving this is by selecting the call cd. Similarly in order for f to learn everything but c we need to select call af. So, until now we have made the LNS- and SPI-permitted call sequence $ab; cb; ed; ef; cd; af$ which leads to

$$(ABDEF, ABC, ABCDE, ABCDE, DEF, ABDEF) \ .$$

- Only the CO- and SPI-permitted call ad will now lead to t. But ad is not LNS-permitted. □

Theorem 5. *There is an* LNS*-reachable distribution that is not* SPI*-reachable.*

Proof. We will show that there is a 16-distribution reachable by LNS but not by SPI. Recall that (ab) represents a call between a and b, which can be instantiated as either ab or ba. Consider the following call sequence $\sigma := \sigma_1; \sigma_2; \sigma_3$, where

$$\begin{aligned} \sigma_1 = \quad & (12); (34); (56); (78); (ab); (cd); (ef); (gh) \\ \sigma_2 = \quad & (23); (45); (67); (81); (bc); (de); (fg); (ha) \\ \sigma_3 = \quad & (1a); (4c); (7h); (6f). \end{aligned}$$

This sequence has three phases σ_1, σ_2, σ_3, as shown on different lines. We can represent this sequence visually as in Figure 2, where the solid lines are calls that happen in σ_1, dashed lines happen in σ_2, and dotted lines in σ_3.

We will show that σ is not SPI-permitted (nor any of its order variants). Suppose towards a contradiction that it is.

In the first stage, the callee member of each pair loses its token. In the second stage, every agent is involved in one more call. If agent 1 still has a token, then 2 does not. So 3 must have a token, otherwise (23) could not take place. But then 4 does not have a token, so 5 must have it, and so on. It follows that in both blocks, either all even agents have lost their token or all odd agents have lost their token (where a, c, e, g are "odd" and b, d, f, h are "even").

Now, consider the third stage. Here, calls $(1a)$, $(4c)$, $(7h)$ and $(6f)$ are supposed to happen. Note that these include every combination of even/odd from both groups: odd number and odd letter $(1a)$, even number and odd letter $(4c)$, odd

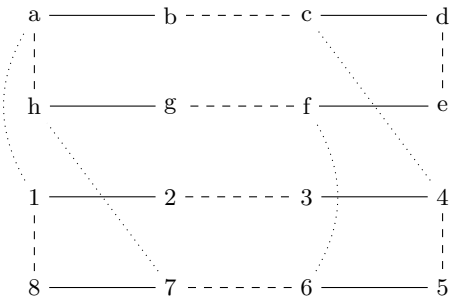

Figure 2. The call sequence $\sigma = \sigma_1; \sigma_2; \sigma_3$.

number and even letter $(7h)$, and even number and even letter $(6f)$. So at least one of these calls is between two agents that do not have their token any more. It follows that the sequence σ is not SPI-permitted.

We still need to show that there is no sequence of other calls that is SPI-permitted and reaches the same distribution of secrets. However, this is fairly straightforward. The distribution s produced by the sequence σ is as follows (let the secret by agent named $i \in \mathbb{N}$ be also i):

$1 : 1278ABGH$	$5 : 3456$	$a : 1278ABGH$	$e : CDEF$
$2 : 1234$	$6 : 5678EFGH$	$b : ABCD$	$f : 5678EFGH$
$3 : 1234$	$7 : 5678ABGH$	$c : 3456ABCD$	$g : EFGH$
$4 : 3456ABCD$	$8 : 1278$	$d : CDEF$	$h : 5678ABGH$

Given this s, we now compute the set of *potential call sequences* $(PC(s), <_s)$. It is easy to show that the set $PC(s)$ consists of all calls in σ plus $(7a)$ and $(6h)$.

Our first observation is that since $A \notin S_6$ and $6 \notin S_a$ it holds that $(67) <_s (7a)$ and that $(7a) <_s (67)$. Thus (67) and $(7a)$ cannot exist in the same call sequence leading to s. It is not difficult to see that (67) is necessary in order to produce s since give the order constraints there is no other way for 6 and 7 to exchange their secrets. Hence, $(7a)$ cannot be used to a call sequence leading to s. In a similar fashion we obtain that $(ha) <_s (6h)$ and that $(6h) <_s (ha)$ and since (ha) is again necessary we conclude that $(6h)$ cannot be used to a call sequence leading to s.

Additional we observe that $(\sigma_1) <_s (\sigma_2)$ and that $(\sigma_2) <_s (\sigma_3)$. Therefore, except for the order of calls within σ_1, σ_2, and σ_3, and the call directions, only σ leads to s. Finally, one can easily verify that σ is also LNS-permitted. □

Theorems 1, 3 and 5 lead to the following corollary. Together with some already discussed inclusions this completes the comparison of the reachability strength between the five protocols (see also Figure 1).

Corollary 1.

1. *There is a* TOK*-reachable distribution that is not* CO*-reachable.*

2. *There is a* TOK*-reachable distribution that is not* SPI*-reachable.*
3. *There is a* CO*-reachable distribution that is not* SPI*-reachable.*

We presented several examples of distributions that are reachable by some of the protocols and unreachable by others. A natural question to ask is "are these distributions optimal counter-examples?", i.e., "did we use the smallest possible number of agents?". We implemented an algorithm that counts the reachable distributions for all five protocols modulo isomorphism (i.e., modulo renaming the agents) [13]. The results of this implementation can be found in Table 1.

Given the inclusions of Theorem 1, Table 1 tells us that all protocols reach the same set of distributions for up to 3 agents. This implies that the 4 distribution in Theorem 3 is optimal. We also see that LNS and CO reach the same set of distributions for up to 5 agents, which implies that the 6-distribution in Theorem 4 is optimal. We do not know whether the non-SPI reachable 16-distribution in the proof of Theorem 5 is optimal (due to a huge combinatorial explosion the implementation in [13] can only count distributions up to at most 7 agents).

Table 1. Number of non-isomorphic reachable distributions for up to 5 agents. For LNS and ANY these numbers are also in the On-Line Encyclopedia of Integer Sequences (OEIS) as `https://oeis.org/A307085` and `https://oeis.org/A318154`, respectively.

n	LNS	CO	SPI	TOK = ANY
2	2	2	2	2
3	4	4	4	4
4	15	15	16	16
5	97	97	111	111

4 Subreachability, Unorderded Distributions

Subreachability in Ordered Distributions While there are distributions that are not even ANY-reachable, all possible distributions are subreachable by any of the five protocols we consider. In [8] this was shown for a more general setting using incomplete network topologies that change dynamically when agents exchange 'phone numbers', but only for the protocol ANY.

Theorem 6. *All distributions are* ANY*-,* CO*-,* LNS*-,* SPI*-,* TOK*-subreachable.*

Proof. We adapt the proof of [8, Section 6.2]. Given a distribution s for agents $\mathcal{A}$, let the *number of secrets known by the agents in* s be defined as $\mathsf{sec}(s) = \Sigma_{a\in\mathcal{A}}|S_a|$, where S_a is the set of secrets known by a in s.

For any protocol P and for any distribution s, we prove by induction on $m = \mathsf{sec}(s)$ that s is P-subreachable. In the base case $m = 1$ the distribution must have shape (A) for a single agent a. This distribution is clearly (sub)reachable by all protocols and the empty call sequence.

Assuming that the result holds for m secrets we will show that it holds for $m+1$ secrets. We need to distinguish two subcases: either there is an agent a who knows a single secret (i.e., an agent who has not made any call yet), or not.

In the first subcase, as $\Sigma_{b\in\mathcal{A}\setminus\{a\}}|S_b| = m$, by induction hypothesis there is a call sequence σ such that the $(\mathcal{A}\setminus\{a\})$-restriction of s is P-subreachable by σ from the initial distribution for the set of agents $\mathcal{A}\setminus\{a\}$. Clearly, s is then P-subreachable by the same call sequence σ from the initial distribution for the set of agents $\mathcal{A}$, as agent a has not been involved in any call. This holds for all five protocols $\mathsf{ANY}, \mathsf{CO}, \mathsf{LNS}, \mathsf{SPI}, \mathsf{TOK}$.

In the second subcase, there must be an agent a who knows at least one other secret B than its own secret A. As $|S_a\setminus\{B\}| + \Sigma_{b\in\mathcal{A},b\neq a}|S_b| = m$, by induction there is a call sequence σ such that s' is P-subreachable by σ, where s' is as s (and defined for the same set of agents) except that $S'_a = S_a\setminus\{B\}$.

First, assume that P is one of ANY, CO, or LNS. Let $c\notin\mathcal{A}$. The role of agent c will be to inform a of B and nothing else. Let s'' be the distribution reached by executing $bc;\sigma;ca$ in the initial distribution for agents $\mathcal{A}\cup\{c\}$. Observe that s is the restriction to $\mathcal{A}$ of s''. Also, call bc is ANY-, CO-, and LNS-permitted, as it is the first call. The last call ca is obviously ANY-permitted. It is CO-permitted because prefix $bc;\sigma$ does not contain a call between c and a. It is also LNS permitted, since c did not learn a in the first call and was not involved in σ. Therefore $bc;\sigma;ca$ is an ANY- CO- and LNS-permitted call sequence reaching s.

Now let $\mathsf{P} = \mathsf{SPI}$. Let in this case $c,d\notin\mathcal{A}$, and consider call sequence $bc;dc;\sigma;da$ for set of agents $\mathcal{A}\cup\{c,d\}$, resulting in distribution s''. In first call bc, b keeps its token, as in the initial distribution for $\mathcal{A}$, but c loses its token (so c can no longer inform a of B at the end, as in the previous case). In the second call dc, d keeps it token and learns B from c. Therefore, in the last call da, d can inform a of B, as desired. Also note that s is the restriction to $\mathcal{A}$ of s''. Therefore $bc;dc;\sigma;da$ is a SPI-permitted call sequence reaching s.

Finally, let $\mathsf{P} = \mathsf{TOK}$. This subcase is fairly similar to the subcase SPI. Again, as for SPI, let $c,d\notin\mathcal{A}$. However, now consider call sequence $cb;dc;\sigma;ca$. In the first call c hands its token to b. So b can still engage in σ as before. In the second call dc agent d hands back a token to agent c. Therefore, the final call ca (instead of da, for SPI) is TOK-permitted resulting in c again informing a of B. Therefore $cb;dc;\sigma;ca$ is a TOK-permitted call sequence reaching s. □

Reachability in Unordered Distributions To illustrate the difference between reachability in unordered and ordered distributions, let us consider the following example. In Theorem 3 we have shown that the ordered distribution $(ABCD, ABCD, ABC, ABD)$ is not CO-reachable. However, this holds only if we understand it as an ordered distribution. It is not difficult to see that the unordered distribution $\{ABCD, ABCD, ABC, ABD\}$ is CO-reachable by the call sequence $ab;ac;bd;cd$.

Theorem 7. *The protocols* ANY, TOK *and* SPI *reach the same unordered distributions.*

Proof. The fact that ANY and TOK reach the same set of ordered distributions (Theorem 2) implies that they also reach the same set of unordered ones. To show that TOK and SPI also reach the same set of unordered distributions we proceed as follows: assume that we have the unordered distribution $\{S_{a_1}, \ldots, S_{a_i}, S_{a_j}, \ldots S_{a_n}\}$ wherein (at least) the agent knowing S_{a_i} possesses a token. Both the TOK and the SPI-call between agents knowing S_{a_i} and S_{a_j} will lead to $\{S_{a_1}, \ldots, S_{a_i} \cup S_{a_j}, S_{a_i} \cup S_{a_j}, \ldots S_{a_n}\}$ where exactly one of the agents that know $S_{a_i} \cup S_{a_j}$ possesses a token. These two unordered distributions are the same, which proves the theorem. □

5 Further Research: Parallel Gossip

We very succinctly describe some results for the setting wherein agents may make calls in parallel. Instead of *sequences of individual calls*, one now considers *sequences of rounds of calls*, where a round of calls consists of a set of calls made in parallel. Different semantics for parallel calls include the 'classical' 1970s setting of gossip [12] wherein calls made in parallel must be mutually disjoint, and the 'modern' 1990s setting of gossip [11] wherein agents, instead, may receive multiple calls. The latter leads to novel reachable distributions, for example, (AB, ABC, BC) is reachable by the simultaneous calls ab, ba, cb, wherein agent b simultaneously receives A from a and C from c. Let us call such a distribution *parallel reachable*, where the notion used so far is *sequential reachable*.

Although $(ABCD, ABCD, ABC, ABD)$ is not sequential CO-reachable (see Theorem 3), it is parallel CO-reachable by the sequence $\{ad, bc, ca, db\}; \{ab\}$ in two rounds. Similarly, $(ABCDEF, ABC, ABCDE, ABCDEF, DEF, ABDEF)$ is not sequential LNS-reachable (Theorem 4), but it is parallel LNS-reachable by the sequence $\{ab, cb, de, fe\}; \{ca, dc, fd, af\}; \{da\}$ in three rounds. Hence reachability in parallel gossip is very different and should be further investigated.

As we mentioned in the introduction the main motivation for studying reachability issues in gossip protocols is to provide an observer with some tools for understanding which protocol is currently being used by the agents. Some further research in this setting could also involve determining the reasoning power that such an observer should have or studying the design of a procedure/determining the resources needed for constructing such observers.

Beyond parallel calls and the observer construction, while in this paper we restricted our attention to only five protocols, our aim is to investigate reachability for protocols that have epistemic conditions. Examples are "call if you know that / consider it possible that an agent will learn a secret" and "don't call if you are an expert". In general, our results should be received as part of a bigger effort to compare the combinatorial properties of epistemic gossip protocols.

Acknowledgements. We would like to thank the anonymous reviewers for their helpful corrections and suggestions. Hans van Ditmarsch is also affiliated to IMSc, Chennai, as associate researcher.

References

1. Apt, K., Grossi, D., van der Hoek, W.: Epistemic protocols for distributed gossiping. In: Proceedings of 15th TARK (2015). https://doi.org/10.4204/EPTCS.215.5
2. Attamah, M., van Ditmarsch, H., Grossi, D., van der Hoek, W.: Knowledge and gossip. In: Proc. of 21st ECAI. pp. 21–26. IOS Press (2014). https://doi.org/10.3233/978-1-61499-419-0-21
3. Attamah, M., van Ditmarsch, H., Grossi, D., van der Hoek, W.: The pleasure of gossip. In: Başkent, C., Moss, L., Ramanujam, R. (eds.) Rohit Parikh on Logic, Language and Society. pp. 145–163. Springer (2017). https://doi.org/10.1007/978-3-319-47843-2_9
4. van Ditmarsch, H., van Eijck, J., Pardo, P., Ramezanian, R., Schwarzentruber, F.: Epistemic protocols for dynamic gossip. J. Applied Logic **20**, 1–31 (2017). https://doi.org/10.1016/j.jal.2016.12.001
5. van Ditmarsch, H., van Eijck, J., Pardo, P., Ramezanian, R., Schwarzentruber, F.: Dynamic gossip. Bulletin of the Iranian Mathematical Society pp. 1–28 (2018). https://doi.org/10.1007/s41980-018-0160-4
6. van Ditmarsch, H., Kokkinis, I., Stockmarr, A.: Reachability and expectation in gossiping. In: An, B., Bazzan, A., Leite, J., Villata, S., van der Torre, L. (eds.) Proceedings of the 20th PRIMA. pp. 93–109. Springer (2017). https://doi.org/10.1007/978-3-319-69131-2_6, lNCS 10621
7. van Ditmarsch, H., Kokkinis, I.: The expected duration of sequential gossiping. In: Belardinelli, F., Argente, E. (eds.) Proceedings of 15th EUMAS. LNCS, vol. 10767, pp. 131–146. Springer (2017). https://doi.org/10.1007/978-3-030-01713-2_10
8. Gattinger, M.: New Directions in Model Checking Dynamic Epistemic Logic. Ph.D. thesis, University of Amsterdam (2018), `https://malv.in/phdthesis`, ILLC Dissertation Series DS-2018-11
9. Göbel, F., Cerdeira, J.O., Veldman, H.J.: Label-connected graphs and the gossip problem. Discrete Mathematics **87**(1), 29–40 (1991). https://doi.org/10.1016/0012-365X(91)90068-D
10. Hedetniemi, S., Hedetniemi, S., Liestman, A.: A survey of gossiping and broadcasting in communication networks. Networks **18**, 319–349 (1988). https://doi.org/10.1002/net.3230180406
11. Kermarrec, A.M., van Steen, M.: Gossiping in distributed systems. SIGOPS Oper. Syst. Rev. **41**(5), 2–7 (2007). https://doi.org/10.1145/1317379.1317381
12. Knödel, W.: New gossips and telephones. Discrete Mathematics **13**, 95 (1975). https://doi.org/10.1016/0012-365X(75)90090-4
13. Kokkinis, I.: Implementation for reachability and expectation in gossiping, `https://github.com/Jannis17/gossip_protocol_expectation`